

УДК 004.9

К. Бєлоусов, Т. Маєвський

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)
(Технічний коледж Тернопільського національного технічного університету імені Івана Пулюя, Україна)

ВЕЛИКІ ДАНІ ТА АНАЛІТИЧНЕ ОПРАЦЮВАННЯ В НАУКОВИХ ДОСЛІДЖЕННЯХ

UDC 004.9

К. Bielousov, T. Maievskyi

BIG DATA AND ANALYTICAL PROCESSING IN SCIENTIFIC RESEARCH

На даний час спостерігається безпрецедентне зростання обсягів колекцій даних накопичених в результаті наукових досліджень. Установи та організації використовують різнотипові аналітичні інструменти, сформовані на основі методів штучного інтелекту (англ. Artificial intelligence, AI) та методів машинного навчання (англ. Machine Learning, ML) для отримання знань та інформації на основі даних, щоб зменшити витрати, збільшити потоки доходів, розробити персоналізовані сервіси та послуги. Невпинно зростає перелік, які використовують «великі дані» (англ. Big Data, BD) для вдосконалення та оптимізації своїх технологічних та бізнес-процесів, виявлення ринкових споживчих тенденцій.

За оцінками фахівців прогнозується, що капіталізація світового ринку «великих даних» до 2025 року перевищить 70 мільярдів доларів США. Водночас очікується, що Сполучені Штати домінуватимуть в цій галузі, сприяючи відбору та аналітичному опрацюванню понад дев'яносто відсотків колективних «великих даних» Північної Америки до 2025 року. Використання «великих даних» в наукових дослідженнях швидко зростає. На даний час обсяги дослідницьких даних, що генеруються за один день, за оцінками фахівців, співмірна з обсягами, які раніше генерувалися за десятиліття. Нещодавні звіти приватних компаній в галузі аналітичного опрацювання даних, в тому числі Всесвітнього економічного форуму, свідчать, що наразі людство оперує сорока чотирма зетабайтами даних. В перспективі ці обсяги зростатимуть до 463 екзабайтів накопичуваних даних щодня в усьому світі [1]. Великі технологічні гіганти, зокрема Google, Facebook, Amazon і Microsoft, зберігають приблизно 1200 петабайт даних завдяки доступу до цифрових технологій по всьому світу. У сукупності статистичні дані свідчать, що розвиток цифровізації, доступність Інтернету та поширення Інтернету речей (IoT) сприяють зростаючому перевантаженню даними.

Складна природа наборів сучасних дослідницьких наборів та колекцій даних формує складну задачу видобування та отримання значущих висновків та знань з «великих даних». Накопичення наборів даних з більш різноманітними колекціями «великих даних» дозволяє ефективніше ідентифікувати загальні закономірності, притаманні загальній масі наборів даних, за рахунок незначних характеристик, які можуть висвітлити важливі області [2]. Тому аналітичне опрацювання «великих даних» повинно використовувати надійні методи для ідентифікації надійних ознак.

Література

1. How Much Data Is Created Every Day? [27 Staggering Stats], 2022. URL: https://earthweb.com/how-much-data-is-created-every-day/?gclid=Cj0KCQIAkMGcBhCSARIsAIW6d0B0cDctoIEGRbWuT3s7MdUY-ppN87LCz9C9Z_RKrkGKvUejqzSVDWEaArPS EALw_wcB.
2. Boehm, Kevin M., et al. «Harnessing multimodal data integration to advance precision oncology.» *Nature Reviews Cancer* 22.2 (2022): 114–126.